CONSISTENT RANKING OF MULTIVARIATE VOLATILITY MODELS

Sébastien Laurent[1], Jeroen V.K. Rombouts[2] and Francesco Violante[3]

December 10, 2008

Abstract

A large number of parameterizations have been proposed to model conditional variance dynamics in a multivariate framework. This paper examines the ranking of multivariate volatility models in terms of their ability to forecast out-of-sample conditional variance matrices. We investigate how sensitive the ranking is to alternative statistical loss functions which evaluate the distance between the true covariance matrix and its forecast. The evaluation of multivariate volatility models requires the use of a proxy for the unobservable volatility matrix which may shift the ranking of the models. Therefore, to preserve this ranking conditions with respect to the choice of the loss function have to be discussed. To do this, we extend the conditions defined in Hansen and Lunde (2006) to the multivariate framework. By invoking norm equivalence we are able to extend the class of loss functions that preserve the true ranking. In a simulation study, we sample data from a continuous time multivariate diffusion process to illustrate the sensitivity of the ranking to different choices of the loss functions and to the quality of the proxy. An application to three foreign exchange rates, where we compare the forecasting performance of 16 multivariate GARCH specifications, is provided.

Keywords: Volatility, Multivariate GARCH, Matrix norm and loss function, Norm equivalence
JEL Classification: C10, C32, C51 C52, C53, G10

---

[1]Université de Namur, CeReFim, B-5000, Namur, Belgium; Université catholique de Louvain, CORE, B-1348, Louvain-la-Neuve, Belgium.

[2]HEC Montréal, CIRANO, CIRPEE and Université catholique de Louvain, CORE, B-1348, Louvain-la-Neuve, Belgium.

[3]Université de Namur, CeReFim, B-5000, Namur, Belgium; Université catholique de Louvain, CORE, B-1348, Louvain-la-Neuve, Belgium.

Financial support from the CREF-HEC Montreal and the Belgian Program on Interuniversity Poles of Attraction initiated by the Belgian State Prime Minister's Office, science policy programming, is gratefully aknowledged. We thank Luc Bauwens, Raouf Boucekkine and Lars Stentoft for helpfull comments.

Correspondence to Sébastien Laurent, FUNDP, Rempart de la Vierge, 8, B-5000 Namur, Belgium.Telephone: +32 81 72 48 69. Fax: +32 81 72 48 40. E-mail: sebastien.laurent@fundp.ac.be

1 Introduction

A special feature of economic forecasting compared to general economic modeling is that we can measure a model's performance by comparing its forecasts to the outcomes when they become available. Generally, several forecasting models are available for the same variable and models are compared through the computation of a loss function. Elliott and Timmermann (2008) provide an excellent survey on the state of the art of forecasting in economics. Details on volatility and correlation forecasting can be found in Andersen, Bollerslev, Christoffersen, and Diebold (2006).

The evaluation of the forecasting performance of volatility models raises an important problem. The variable of interest (i.e. volatility) being unobservable, the evaluation of the loss function has to rely on a proxy which may change the ordering. The impact on the ordering of the substitution of the true volatility by a proxy has been investigated in detail by Hansen and Lunde (2006). They provide conditions, for both the loss function and the volatility proxy, under which the approximated ranking (based on the proxy) is consistent for the true ranking (based on the true, but unobservable volatility).

Hansen and Lunde's (2006) results have important implications for all testing procedures for superior predictive ability as in Diebold and Mariano (1995), West (1996), Clark and McCracken (2001), the reality check by White (2000), and the recent contributions of Hansen and Lunde (2005) with the superior predictive ability (SPA) test and Hansen, Lunde, and Nason (2003) with the Model Confidence Set test, among others. When the target variable is unobservable, an unfortunate choice of the loss function may deliver unintended results even when the testing procedure is formally valid.

In this paper, we extend findings of Hansen and Lunde (2006) to the multivariate framework. To forecast the conditional variance matrix of a portfolio of financial assets, we focus on multivariate GARCH models (MGARCH) (see Bauwens, Laurent, and Rombouts (2006) for a survey), though the extension to other multivariate volatility models, like stochastic volatility and Markov switching models is straightforward. With respect to ranking models in the multivariate GARCH class, where conditional variance matrices are compared, little is known about the properties of the loss function.

We have four main contributions in this paper. First, we select a set of loss functions well suited to evaluate the differences in sequences of symetric positive definite matrices. We consider six different loss functions based on matrix norms, namely the p-norm with $p = 1$ and $p = 2$ (the latter is known as the Frobenius norm), the spectral norm and their squared transformations. These loss functions are frequently used in practice.

Second, we derive conditions for consistent ranking for the multivariate case and we show that, though violating their conditions, a loss function might still lead to a consistent ranking if norm equivalence can be invoked with respect to a consistent loss function. This result allows us to extend the requirements stated for univariate volatility models in Hansen and Lunde (2006) allowing loss functions that would be classified as inconsistent in the univariate case. For each of the loss functions considered in this paper we verify whether they satisfy the conditions to ensure a consistent ranking.

Third, through a comprehensive Monte Carlo simulation, we study the impact of the deterioration of the quality of the proxy on the ranking of MGARCH models with respect to different choices for the loss function. The true model is a multivariate diffusion from which we compute the integrated covariance as the true daily covariance. The MGARCH models are estimated on daily returns and used to compute 1-step ahead forecasts. The proxy of the daily covariance is Realized Covariance as defined in Andersen, Bollerslev, Diebold, and Labys (2003). The quality of this proxy is controlled through the level of aggregation of the simulated intraday data used to compute Realized Covariance. The main conclusion of this simulation is that inconsistent loss functions are not *per se* inferior to consistent ones. When the quality of the proxy is sufficiently good, consistency between the true and the approximated ranking can still be achieved. As the accuracy of the proxy deteriorates, the objective bias (i.e. the discrepancy between the true and the approximated ranking) becomes relevant and may affect the ordering between models.

Fourth, we illustrate our findings using three exchange rates (Euro, UK pound and Japanese yen against US dollar). We consider 16 MGARCH specifications which are frequently used in practice. The advantage of choosing a consistent loss function to evaluate model performances is striking. The ranking based on an inconsistent loss function, together with an uninformative proxy, is found to be severely biased. In fact, inferior models, that is models based on the RiskMetrics approach, emerge though it is unlikely that these are the best forecasting models. Overall, the set of 16 MGARCH models seem to produce conditional variance matrix predictions that are quite close.

The rest of the paper is organized as follows. Section 2 introduces the set of selected loss functions and revisits Hansen and Lunde's (2006) conditions for consistent ranking. An additional condition, based on the notion of norm equivalence, is introduced. Section 3 provides a brief overview of several GARCH specification considered in this paper and thus constituting the forecasting models set. In Section 4, we introduce realized covariance as a proxy for the unobserved conditional variance matrix. A detailed simulation study in Section 5 investigates the robustness of the ranking subject to consistent and inconsistent loss functions with respect to the level of accuracy of the proxy. The empirical application is presented in Section 6. Section 7 concludes.

2 Consistent ranking and distance metrics

As explained in Andersen, Bollerslev, Christoffersen, and Diebold (2006), the problem when comparing and ranking forecasting performance of volatility models is that the true conditional variance is unobservable so that a proxy for it is required. Let us define the true, or underlying, ordering between volatility models as the ranking implied by a loss function, evaluated with respect to the unobservable conditional covariance. The substitution of the latter by a proxy may introduce, because of its randomness, a ranking of volatility models that differs from the true one. Hansen and Lunde (2006) provide a theoretical framework for the analysis of the ordering of stochastic sequences and identify conditions that a loss function and the volatility proxy have to satisfy to deliver an ordering consistent with the true ranking when a proxy for the conditional covariance is used. In this section, we discuss and extend these conditions to the case of multivariate volatility models.

We first fix some basic notations and make explicit what we mean by consistent ranking. For N time series at time t we have M candidate models for the conditional variance matrix denoted by H_{it} $i = 1, \ldots, M$. Define $L(\cdot, \cdot)$ an integrable loss function from $R^{N \times N} \rightarrow R^{+}$ such that $L(\Sigma_t, H_{it})$ is the loss function using the true but unobservable conditional variance matrix Σ_t. Similarly $L(\hat{\Sigma}_t, H_{it})$ is the loss function using $\hat{\Sigma}_t$, a proxy of Σ_t. Consistency of ranking means that $E(L(\Sigma_t, H_{it})) \geq E(L(\Sigma_t, H_{jt})) \Leftrightarrow E(L(\hat{\Sigma}_t, H_{it})) \geq E(L(\hat{\Sigma}_t, H_{jt}))$ is true for all $i \neq j$.

2.1 Hansen and Lunde's (2006) conditions for consistent ranking

Without loss of generality we can redefine the function $L(.,.)$ from the space of the $N \times N$ matrices to R^{+} as a scalar valued function from $R^{N(N+1)/2} \rightarrow R^{+}$ of all unique elements of the matrices Σ_t and H_{it} since these are covariance matrices and therefore symetric. Let us denote $\sigma_t = [\sigma_{kj,t}] = vech(\Sigma_t)$ and $h_{i,t} = [h_{kj,it}] = vech(H_{it})$ where $vech(.)$ is the operator that stacks the lower triangular portion of a matrix into a vector. As developed by Hansen and Lunde (2006) for univariate volatility models, similar relevant sufficient conditions to achieve consistency of the ranking for multivariate models are:

(i) $L(\Sigma_t, H_{it})$ and $L(\hat{\Sigma}_t, H_{it})$ have the same parametric form $\forall i$ so that uncertainty depends only on $\hat{\Sigma}_t$ and $\Im_t$ is a filtration such that Σ_t and H_{it} are $\Im_{t-1}$ measurable.

(ii) $\frac{\partial^2 L(\Sigma_t, H_{it})}{\partial \sigma_{kj,t} \partial \sigma_{kj,t}}$ is finite and does not depend on $h_{kj,it}$ $\forall k, j$, and, $\xi_t = (\hat{\sigma}_t - \sigma_t)$ is a vector martingale difference sequence with respect to $\Im_t$.

To illustrate the validity of the above conditions, consider the second order Taylor expansion of $L(\Sigma_t, H_{it})$ around the true value Σ_t:

$$L(\hat{\Sigma}_t, H_{it}) \cong L(\Sigma_t, H_{it}) + \left(\frac{\partial L(\Sigma_t, H_{it})}{\partial \sigma_t} \right)' (\hat{\sigma}_t - \sigma_t) + \frac{1}{2} \left[(\hat{\sigma}_t - \sigma_t)' \frac{\partial^2 L(\Sigma_t, H_{it})}{\partial \sigma_t \partial \sigma_t'} (\hat{\sigma}_t - \sigma_t) \right]. \quad (1)$$

Taking conditional expections with respect to $\Im_{t-1}$ we get

$$E(L(\hat{\Sigma}_t, H_{it})|\Im_{t-1}) \quad \cong \quad E(L(\Sigma_t, H_{it})|\Im_{t-1}) + \frac{1}{2}\left[E\left(\xi_t^{'}\frac{\partial^2 L(\Sigma_t, H_{it})}{\partial\sigma_t\partial\sigma_t^{'}}\xi_t|\Im_{t-1}\right)\right]. \qquad (2)$$

When condition (i) and (ii) are satisfied we have

(a) $E((L^{'}(\Sigma_t, H_{it}))^{'}\xi_t|\Im_{t-1}) = (L^{'}(\Sigma_t, H_{it}))^{'}E(\xi_t|\Im_{t-1}) = 0$ whenever $\hat{\sigma}_t$ is conditionally unbiased with respect to σ_t;

(b) $\left(\frac{\partial^2 L(\Sigma_t, H_{it})}{\partial\sigma_t\partial\sigma_t}\right) = f(\sigma_t^2, .)$ does not depend on model i.

Hence $E(L(\hat{\Sigma}_t, H_{it})|\Im_{t-1})$ and $E(L(\Sigma_t, H_{it})|\Im_{t-1})$ induce the same ordering over i.

The discrepancy between the true and the approximated ordering which is likely to occur whenever condition (ii) is violated, is defined as objective bias. The objective bias must not be confused with the sampling error. While the latter tend to disappear asymptotically (i.e. $T^{-1}\Sigma_t L(\hat{\Sigma}_t, H_{it}) \xrightarrow{p} E(L(\hat{\Sigma}_t, H_{it})))$, the presence of the objective bias may induce the sample evaluation to be inconsistent for the true one independently from the sample size.

To conclude, (2) implies that in order to achieve consistency of the approximated ranking, the equivalence between $E(L(\hat{\Sigma}_t, H_{it})|\Im_{t-1})$ and $E(L(\Sigma_t, H_{it})|\Im_{t-1})$ is not required, but it is sufficient that the discrepancy, $\left[E\left(\xi_t^{'}f(\sigma_t^2, .)\xi_t|\Im_{t-1}\right)\right]$ is constant across models, thus not affecting the ranking. Notice, that the last term on the right hand side of (2) depends on the variance of the proxy. Hence, even if condition (ii) is violated, that is $\left(\frac{\partial^2 L(\Sigma_t, H_{it})}{\partial\sigma_t\partial\sigma_t}\right) = f(\sigma_t^2, h_{it})$, the more accurate the proxy, the less likely the objective bias. That is, $\left[E\left(\xi_t^{'}f(\sigma_t^2, h_{it})\xi_t|\Im_{t-1}\right)\right]$, though depending on i, becomes negligible, leaving the ranking unaffected.

2.2 Norm equivalence

When the loss function is defined in terms of a matrix norm on the space of $N \times N$ positive definite matrices, $S^{N\times N}$, that is $L(\Sigma_t, H_{it}) = \|.\|_a$, a useful property of matrix norms, namely the norm equivalence, can be invoked. Norm equivalence is defined as follows (see Golub and Van Loan, 1996 or Horn and Johnson, 1985 for details). For any two matrix norms $\|.\|_a$ and $\|.\|_b$ on a finite dimensional space, norm equivalence is defined as

$$k\,\|A\|_a \leq \|A\|_b \leq l\,\|A\|_a\,, \qquad (3)$$

for $0 < k < l < \infty$ and $A \in S^{N\times N}$. If two norms are equivalent then they introduce the same topology on $S^{N\times N}$. This property is preserved under functional transformations - e.g. $f(.)$ and $g(.)$ of the matrix norms $\|.\|_a$ and $\|.\|_b$, provided $f(\|.\|_a)$ and $g(\|.\|_b)$ have the same degree of homogeneity.

A loss function based on matrix norms that satisfies condition (i) but violates (ii) may still yield a consistent ordering. This is the case when norm equivalence between a consistent and an inconsistent loss function can be established. In this case, the inconsistent loss function can introduce the same ordering as the consistent one. This allows for the additional condition (iii) which can be stated as follows.

(iii) Given a consistent loss function $L_a(\Sigma_t, H_{it})$ and another loss function $L_b(\Sigma_t, H_{it})$: $kL_a(\Sigma_t, H_{it}) \leq L_b(\Sigma_t, H_{it}) \leq lL_a(\Sigma_t, H_{it})$ holds for $0 < k < l < \infty$.

If (i) and (iii) holds for $L_b(\Sigma_t, H_{it})$ then it is equivalent to the consistent loss function $L_a(\Sigma_t, H_{it})$ and thus induces the same ordering.

The next section focuses on two families of loss functions which have been widely used in the literature: the p-norm, the eigenvalue norm, and we also consider their square transformation. It should be noted that while the p-norm and the Eigenvalue norm are valid norms, their squared versions are not since, though sharing most of the properties of matrix norms, they violate the homogeneity assumption, as they are homogeneous of degree two.

2.3 P-norm loss function

The p-norm between two matrices (Σ_t and H_{it}) is defined as the p^{th}-root of the sum of element-wise differences to the power p, i.e.

$$L(\Sigma_t, H_{it})_p = \left(\sum_{1 \leq k,j \leq N} |\sigma_{kj,t} - h_{kj,it}|^p \right)^{1/p}. \tag{4}$$

When $p = 2$ this norm is known as the Frobenius norm, while $L(\Sigma_t, H_{it})_2^2$ denotes its square transformation. It is easy to show that, while the Frobenius norm does not satisfy condition (ii), its squared version does, provided that $\hat{\Sigma}_t$ is conditionally unbiased for Σ_t, that is when $E(\hat{\sigma}^2_{kj,t}|\Im_{t-1}) = \sigma^2_{kj,t}$. In this paper, we consider also $p = 1$ (sum of absolute element-wise differences) and its square $L(\Sigma_t, H_{it})_1^2$. The p-norm with $p = 1$ (and its square) violates condition (ii) because it is not differentiable. However, we can show that $L(\Sigma_t, H_{it})_1^2$ satisfies condition (iii) with respect to $L(\Sigma_t, H_{it})_2^2$ because of the following inequalities:

$$L(\Sigma_t, H_{it})_2^2 \leq L(\Sigma_t, H_{it})_1^2 \leq N^2 L(\Sigma_t, H_{it})_2^2, \tag{5}$$

which comes directly from the norm equivalence between $L(\Sigma_t, H_{it})_2$ and $L(\Sigma_t, H_{it})_1$. We illustrate the proof in the bivariate case ($N = 2$). Let $a_{kj,t} = (\sigma_{kj,t} - h_{kj,it})$ $\forall k, j = 1, 2$, so that

$$L(\Sigma_t, H_{it})_2^2 = \sum_{k,j=1,2} a^2_{kj,t} \tag{6}$$

$$L(\Sigma_t, H_{it})_1^2 = \left[\sum_{k,j=1,2} |a_{kj,t}| \right]^2 \tag{7}$$

$$= \sum_{k,j=1,2} a^2_{kj,t} + 2|a_{12,t}|^2 + 2|a_{11,t}||a_{22,t}| + 4|a_{11,t}||a_{12,t}| + 4|a_{12,t}||a_{22,t}| \tag{8}$$

$$\geq \quad L(\Sigma_t, H_{it})_2^2, \tag{9}$$

since for any two positive scalars a_{kj} and a_{lm}, $2a_{kj}a_{lm} \leq a_{kj}^2 + a_{lm}^2$. We also have that

$$L(\Sigma_t, H_{it})_1^2 \leq 4 \sum_{k,j=1,2} a_{kj,t}^2 = 4L(\Sigma_t, H_{it})_2^2, \tag{10}$$

which proves the result in (5). Using similar arguments for the p-norm with $p = 1$ we have

$$L(\Sigma_t, H_{it})_2 \leq L(\Sigma_t, H_{it})_1 \leq NL(\Sigma_t, H_{it})_2. \tag{11}$$

But in this case, since both $L(\Sigma_t, H_{it})_1$ and $NL(\Sigma_t, H_{it})_2$ do not satisfy condition (ii), though equivalent, condition (iii) cannot be applied because condition (iii) is violated.

2.4 Eigenvalue loss function

The eigenvalue norm, also called spectral norm, is widely used in principal component analysis. It is defined as the square root of the largest eigenvalue of the matrix $(\Sigma_t - H_{it})^2$ and denoted by $L(\Sigma_t, H_{it})_E = \sqrt{\lambda_{max}(\Sigma_t, H_{it})}$. As before, we also consider its square transformation, i.e. $L(\Sigma_t, H_{it})_E^2 = \lambda_{max}(\Sigma_t, H_{it})$. As an illustration, the square of the eigenvalue norm becomes in the bivariate case

$$\begin{aligned} L(\Sigma_t, H_{it})_E^2 &= \lambda_{\max}[(\Sigma_t - H_{it})^2] && (12) \\ &= \frac{1}{2} f(\sigma_{11,t}, \sigma_{12,t}, \sigma_{22,t}, h_{ij,t}) + \frac{1}{2}\sqrt{g(\sigma_{11,t}, \sigma_{12,t}, \sigma_{22,t}, h_{ij,t})}, && (13) \end{aligned}$$

where

$$\begin{aligned} f(\sigma_{11,t}, \sigma_{12,t}, \sigma_{22,t}, h_{kj,it}) &= (\sigma_{11,t} - h_{11,it})^2 + 2(\sigma_{12,t} - h_{12,it})^2 + (\sigma_{22,t} - h_{22,it})^2 && (14) \\ g(\sigma_{11,t}, \sigma_{12,t}, \sigma_{22,t}, h_{kj,it}) &= \sqrt{\begin{array}{c} \left[(\sigma_{11,t} - h_{11,it})^2 - (\sigma_{22,t} - h_{22,it})^2\right]^2 + \\ 4(\sigma_{12,t} - h_{12,it})^2 \left[(\sigma_{11,t} - h_{11,it}) + (\sigma_{22,t} - h_{22,it})\right]^2 \end{array}}\,. && (15) \end{aligned}$$

The second derivative of the loss function with respect to $\sigma_{kj,t}^2$ is

$$\frac{1}{2}\left(f''_{\sigma_{kj,t}^2} + g''_{\sigma_{kj,t}^2}\right). \tag{16}$$

Since $g'_{\sigma_{kj,t}^2}$ and $g''_{\sigma_{kj,t}^2}$ depend on $h_{kj,t}$ condition (ii) is violated. However, we can show that condition (iii) is satisfied with respect to the square of the Frobenius norm which in turn is consistent by condition (ii). We can rewrite the Frobenius norm as

$$L(\Sigma_t, H_{it})_2^2 = Trace[(\Sigma_t - H_{it})^2] = \sum_N \lambda_i, \tag{17}$$

where λ_i are the postive eigenvalues of the matrix $(\Sigma_t - H_{it})^2$. Therefore, we have

$$L(\Sigma_t, H_{it})_E^2 = \lambda_{max} \quad \leq \quad \sum_N \lambda_i = L(\Sigma_t, H_{it})_2 \tag{18}$$

$$L(\Sigma_t, H_{it})_E^2 = \lambda_{max} \quad \geq \quad \bar{\lambda} = N^{-1} \sum_N \lambda_i = N^{-1} L(\Sigma_t, H_{it})_2, \tag{19}$$

which proves the following equivalence:

$$N^{-1} L(\Sigma_t, H_{it})_2^2 \leq L(\Sigma_t, H_{it})_E^2 \leq L(\Sigma_t, H_{it})_2^2. \tag{20}$$

Therefore, using this loss function yields a consistent ranking because $L(\Sigma_t, H_{it})_2^2$ does.

As explained in Section 2.3, if we consider the spectral norm itself (i.e. the square root of the highest eigenvalue of the matrix $(\Sigma_t - H_{it})$) then by norm equivalence it holds

$$N^{-1/2} L(\Sigma_t, H_{it})_2 \leq L(\Sigma_t, H_{it})_E \leq L(\Sigma_t, H_{it})_2. \tag{21}$$

This confirms $L(\Sigma_t, H_{it})_E$ to be ranking inconsistent because condition (iii) is violated.

3 Forecasting models set

In this paper, we focus on the ranking of multivariate volatility models that belong to the MGARCH class. Consider a N-dimensional discrete time vector stochastic process r_t. Let $\mu_t = E(r_t|\Im_{t-1})$ be the conditional mean vector and $H_{it} = E(r_t r_t'|\Im_{t-1})$ the conditional variance matrix for specification i so that we can write the model of interest as:

$$r_t \quad = \quad \mu_t + H_{it}^{1/2} z_t, \tag{22}$$

where $H_{it}^{1/2}$ is a $(N \times N)$ positive definite matrix and z_t is an idependent and identically distributed random innovation vector with $E(z_t) = 0$ and $Var(z_t) = I_N$.

In the application, we consider 16 specifications for H_{it} which are frequently used in practice. For the simulation study, we take a slightly different forecasting models set made up of 10 models, details are in Section 5, in order to control for the degree of similarity between models. The specifications considered in this paper are: the diagonal BEKK of Engle and Kroner (1995) and the multivariate RiskMetrics procedure, J.P.Morgan (1996), developed by J.P. Morgan. The set also includes four variations of the Constant Correlation (CCC) model (Bollerslev, 1990), of the Dynamic Conditional Correlation (DCC) model of Engle (2002), and of the Generalized Orthogonal GARCH (GOGARCH) model of van der Weide (2002), with GARCH (Bollerslev, 1986), GJR (Glosten, Jagannathan, and Runkle, 1992), Exponencial GARCH (Nelson, 1991), Asymmetric Power ARCH (Ding, Granger, and Engle, 1993), Integrated GARCH (Engle and Bollerslev, 1986), RiskMetrics (J.P.Morgan,1996) and Hyperbolic GARCH (Davidson, 2004) specifications for the conditional variance equations. In the GJR model, the impact of squared innovations on the conditional variance is different when the innovation is positive or negative. The asymmetric power ARCH model (APARCH) is a general specification which includes seven other ARCH extensions as special cases. The Exponential GARCH model (EGARCH) accomodates the asymmetric relation between shocks and volatility by expressing the latter as a function of both the magnitude and the sign of the shock. The Integrated GARCH (IGARCH) model is a variation of the GARCH model in which the sum of the variance parameters are constrained to be equal to one, while the RiskMetrics model (RM) is basically an IGARCH model where the constant is set to zero and the ARCH and GARCH coefficients are fixed ex ante. Finally, the Hyperbolic GARCH model (HYGARCH) allows to account for long run dependence in the volatility. The functional forms for H_t are briefly defined in Table 1. See Bauwens, Laurent, and Rombouts (2006) for further details. All the specifications are characterized by a constant conditional mean and the models are estimated by quasi maximum likelihood using G@RCH 5.0 (Laurent, 2007). The sample log-likelihood is given by (up to a constant)

$$-\frac{1}{2}\sum_{t=1}^{T}\log \mid H_{it} \mid -\frac{1}{2}\sum_{t=1}^{T}(r_t-\mu)' H_{it}^{-1}(r_t-\mu), \tag{23}$$

and we maximize numerically for μ and the parameters in H_t.

4 A proxy for the conditional variance matrix

An interesting aspect of volatility is that it becomes observable ex-post. Recent literature has focused on defining a theoretical framework for the estimation of the conditional variance of financial assets returns, which is essentially based on the analysis of high frequency data. McAller and Medeiros (2008) provide a survey on this subject. Following Andersen, Bollerslev, Diebold, and Labys (2003), we rely on the realized covariance ($RCov$) to proxy the ex post variance. In the ideal case of no microstructure noise, this measure, being based on intraday observations, is characterized by a degree of accuracy that decreases as sampling frequency lowers.

Let us assume the observed return vector to be generated by a conditionally normal N-dimensional log-price diffusion $dy(u)$ and a $(N(N+1)/2)$-dimensional covariance diffusion $d\sigma(u)$, with $\sigma(u) = vech(\Sigma(u)) = [\sigma_{ij}(u)]$ for $i,j = 1,...,N$, $i \geq j$ and $u \in [t,t+1]$, with mean vector process $b(u)du$ and covariance matrix $a(u) = s(u)s(u)'$, driven by a $N(N+3)/2$ vector of independent standard Brownian motions $W(u)$. Hence the diffusion process of the system admits the following representation

$$\begin{bmatrix} dy(u) \\ d\sigma(u) \end{bmatrix} = b(u)du + s(u)dW(u), \tag{24}$$

with $b(u)$ and $s(u)$ locally bounded and measurable. Consider now the following partition for the covariance matrix of the system in (24) as

$$a(u) = s(u)s(u)' = \begin{bmatrix} \Sigma(u) & Cov(dy_u, d\sigma_u) \\ Cov(dy_u, d\sigma_u) & Var(d\sigma_u) \end{bmatrix}. \tag{25}$$

Table 1: Summary of the forecasting models set

Model	H_t	# parameters in MGARCH(1,1) $N = 2$
DBEKK (diagonal BEKK)	$H_t = C_0^{*'} C_0^* + A^{*'} \epsilon_{t-1}\epsilon'_{t-1} A^* + G^{*'} H_{t-1} G^*$	$\frac{N(N+5)}{2}$ 7
RiskMetrics	$H_t = (1-\alpha)\epsilon_{t-1}\epsilon'_{t-1} + \alpha H_{t-1}$ $\alpha = 0.96$	0
GOGARCH (generalized orthogonal GARCH)	$H_t = V^{1/2} Z Q_t Z V^{1/2}$, $Q_t = diag(\sigma^2_{p_{1,t}}, \ldots, \sigma^2_{p_{m,t}})$ $Z = P\Lambda^{1/2} U$, $U = \prod_{i<j} R_{i,j}(\delta_{i,j})$ $\quad -\pi \leq \delta_{i,j} \leq \pi$	1 + univ. GARCH
DCC (dynamic conditional correlations)	$H_t = D_t R_t D_t$ $R_t = Q_t^{*-1} Q_t Q_t^{*-1}$ $D_t = diag(h_{11t}^{1/2} \ldots h_{NNt}^{1/2})$ $Q_t = (1-\alpha-\beta)\bar{Q} + \alpha u_{t-1} u'_{t-1} + \beta Q_{t-1}$ $Q_t^{*-1} = diag(q_{11t}^{1/2} \ldots q_{NNt}^{1/2})$	3 + univ. GARCH
CCC (constant conditional correlations)	$H_t = D_t R D_t$ $D_t = diag(h_{11t}^{1/2} \ldots h_{NNt}^{1/2})$	1 + univ. GARCH
	Univariate GARCH models in Q_t and D_t	
$GARCH(1,1)$	$h_{l,t} = \omega_l + \alpha_l \epsilon^2_{l,t-1} + \beta_l h_{l,t-1}$	3 $\forall l = 1, \ldots, N$
$EGARCH(1,0)$	$log(h_{l,t}) = \omega_l + g(z_{l,t-1}) + \beta_l log(h_{l,t-1})$ $g(z_{l,t-1}) = \theta_{l,1} z_{l,t-1} + \theta_{l,2}(\lvert z_{l,t}\rvert - E(\lvert z_{l,t}\rvert))$	4 $\forall l = 1, \ldots, N$
$GJR(1,1)$	$h_{l,t} = \omega_l + \alpha_l \epsilon^2_{l,t-1} + \gamma_l S^-_{l,t-1} \epsilon^2_{l,t-1} + \beta_l h_{l,t-1}$ $S^-_{l,t} = 1$ if $\epsilon_{l,t} < 0$; $S^-_{l,t} = 0$ if $\epsilon_{l,t} \geq 0$	4 $\forall l = 1, \ldots, N$
$APARCH(1,1)$	$h^{\delta_l}_{l,t} = \omega_l + \alpha_l [\lvert\epsilon_{l,t-1}\rvert - \gamma_l \epsilon_{l,t-1}]^{\delta_l} + \beta_l h^{\delta_l}_{l,t-1}$	5 $\forall l = 1, \ldots, N$
$HYGARCH(1,d,1)$	$h_{l,t} = \omega_l [1-\beta_l]^{-1} + \left\{1 - [1-\beta_l]^{-1} \alpha_l [1 + \gamma_l (1-L)^d]\right\} \epsilon^2_{l,t}$	5 $\forall l = 1, \ldots, N$

Since Σ_u identifies the continuous time process for the covariance matrix of the returns, we can define the Integrated Covariance (ICov) as (see Barndorff-Nielsen and Shephard, 2004)

$$ICov_{t+1} = \int_t^{t+1} \Sigma(u)du. \tag{26}$$

Let us now define the intraday returns as $r_{t+\Delta} = y_{t+\Delta} - y_t$ for $t = \Delta, 2\Delta, ..., T$ and $\Delta = 1/m$, where m is the number of intervals per day. In this setting $ICov_t$ can be consistently estimated by the Realized Covariance ($RCov$) (Andersen, Bollerslev, Diebold, and Labys, 2003) which is defined as

$$RCov_{t+1,\Delta} = \sum_{i=1}^{1/\Delta} r_{t+i\Delta} r'_{t+i\Delta}. \tag{27}$$

In fact, since the process defined by (24) does not allow for jumps in the returns, it holds that

$$\underset{\Delta \to 0}{\text{plim}} RCov_{t+1,\Delta} = ICov_{t+1}. \tag{28}$$

In this paper, the $RCov$ serves as a proxy for the true conditional variance matrix when evaluating the forecasting performance of the different MGARCH models. The result (28) suggests that the higher is the intraday frequency used to compute $RCov$, and hence the amount of information available, the higher the accuracy of the proxy.

However, as noted by Andersen, Bollerslev, Diebold, and Labys (2003), positive definiteness of the covariance matrix is ensured only if the number of assets is larger then m (where m is the number of intervals per day). When this condition is violated then the realized covariance matrix fails to be of full rank (i.e. $rank(Rcov) = m < dim(RCov)$) and $RCov$ will meet only the weaker requirement to be semi-positive definite.

5 Simulation study

We investigate the ranking of the MGARCH models with respect to two main dimensions: the quality of the volatility proxy and the choice of the loss function. As expected, we find that if the quality of the proxy is good, both consistent and inconsistent loss functions rank properly. However, when the quality of the proxy is poor, only the consistent loss functions rank properly. Our findings also hold when the sample size in the estimation period increases.

5.1 Setup

Varying the quality of the proxy requires the simulation of a multivariate diffusion process. For our simulation, we select the bivariate CCC-EGARCH(1,0) model (see Table 1) which admits a diffusion limit of the type introduced by (24), defined by the continuous time vector stochastic process $[y_{1t}, y_{2t}, log(\sigma_{1t}^2), log(\sigma_{2t}^2)]'$, with drift and scale given respectively by

$$b(y, \Sigma) = \begin{bmatrix} 0 \\ 0 \\ \omega_1 - \theta_1 \log(\sigma_{1t}^2) \\ \omega_2 - \theta_2 \log(\sigma_{2t}^2) \end{bmatrix} \tag{29}$$

and

$$\begin{aligned} a(y, \Sigma) &= s(y, \Sigma)s(y, \Sigma)' \\ &= \begin{bmatrix} \sigma_{1t}^2 & \rho\sigma_{1t}\sigma_{2t} & \alpha_1\sigma_{1t} & \rho\alpha_2\sigma_{1t} \\ \rho\sigma_{1t}\sigma_{2t} & \sigma_{2t}^2 & \rho\alpha_1\sigma_{2t} & \alpha_2\sigma_{2t} \\ \alpha_1\sigma_{1t} & \rho\alpha_1\sigma_{2t} & \alpha_1^2 + \gamma_1^2(1 - 2/\pi) & \rho\alpha_1\alpha_2 + \gamma_1\gamma_2 C \\ \rho\alpha_2\sigma_{1t} & \alpha_2\sigma_{2t} & \rho\alpha_1\alpha_2 + \gamma_1\gamma_2 C & \alpha_2^2 + \gamma_2^2(1 - 2/\pi) \end{bmatrix}, \end{aligned} \tag{30}$$

where $C = \frac{2}{\pi}\left[\sqrt{1-\rho^2} + \rho \arcsin(\rho) - 1\right]$. The conditional covariance is computed, at each point in time as $\sigma_{(1,2),t} = \rho\sqrt{\sigma_{1,t}^2\sigma_{2,t}^2}$. The matrix $s(y, \Sigma)$ is computed from $a(y, \Sigma)$ by spectral decomposition.

The CCC-EGARCH specification has been preferred to alternative MGARCH specifications - e.g. the DCC model - because it is sufficient to ensure a certain degree of dissimilarity between the true DGP and the set of models while keeping the limiting diffusion fairly tractable.

For the simulation study we use the following parameter values: $\omega_i = -0.02$, $\theta_i = 1-\beta_i = 0.03$, $\alpha_i = -0.09$, $\gamma_i = 0.4$ and $\rho = 0.9$. Our results are based on 500 replications with an estimation sample of 2000 observations and a forecasting sample of T=500 observations. The continuous time process of (24) is approximated by generating 7200 observations per day - i.e. 5 observations per minute. The set of MGARCH models is estimated on daily returns and recursive 1-step ahead forecasts are computed.

The true conditional covariance matrix is measured by the integrated covariance ($ICov$) defined in (26). To proxy the daily covariance matrix, we use the realized covariance ($RCov_{t,\triangle}$), as defined in (27), based on intraday returns sampled at 14 different frequencies, ranging from 1 minute (most accurate) to 24 hour (least accurate), over the forecasting horizon. It is important to stress that given the bivariate DGP we should in principle stop at the 12 hour frequency to ensure a positive definite realized variance matrix at time t. Though, when reporting our simulation results next, we also include the 24 hour frequency to investigate what happens when a realized variance matrix which is not positive definite enters the loss functions.

As underlined by Hafner (2007) it is difficult to derive temporal aggregation results for the process generated by (24) and (29)-(30) due to the non-linearity of the variance matrix Σ_t. The only thing which we require, in the spirit of Meddahi (2002) and Voev and Lunde (2006), is to generate continuous time paths such that the resulting $RCov$ estimators, at different time sampling frequencies, are consistent for $ICov$. Contrary to the previous literature, the diffusion approximation we introduce here allows to control better for the nature and the size of the leverage effect and to preserve the correlation structure of the vector stochastic process $[y_{1t}, ..., \sigma^2_{1t}, ...]'$ ensuring internal consistency of the model.

Note that since we are comparing estimated models, the underlying order, other than for the best model, is unknown. The initial set of models is defined such that all models are expected to be inferior. Apart from the CCC-EGARCH(1,0), the set of models includes the diagonal BEKK, RiskMetrics, CCC-GARCH(1,1), CCC-IGARCH(1,1), CCC-RiskMetrics, GOGARCH-GARCH(1,1), GOGARCH-EGARCH(1,0), GOGARCH-IGARCH(1,1) and GOGARCH-HYGARCH(1,1).

The considered loss functions and their classification are summarized in Table 2.

Table 2: Classification of the loss functions

Matrix Norms (MN)	Type	Transform. of MN	Type
p-norm (p=1)	inconsistent	p-norm (p=1) squared	consistent
Frobenius norm	inconsistent	Frobenius norm squared	consistent
Eigenvalue norm	inconsistent	Eigenvalue norm squared	consistent

5.2 Sample performance ranking and objective bias

We focus first on the ability of the loss function to detect the true model as the best. We compute the frequencies at which each model shows the smallest sample performance where the latter is defined as the mean value of the loss function $T^{-1}\sum_T L(\Sigma_t, H_{it})$.

Table 3 reports these frequencies for the consistent loss functions: the squares of the Frobenius norm, the p-norm with $p = 1$ and of the eigenvalue norm. Unsurprisingly, we find the CCC-EGARCH model ranking first most often for all consistent loss functions at all frequencies for $RCov$. When $ICov$ is used, this frequency is about 51%. The remaining 49% is distributed among the other models (from 0% to 7%) in such a way that no model dominates. One exception is the GOGARCH-EGARCH (17%) which is not surprising since this model is the only one in the set that allows for a leverage effect. Note that the frequency associated to the GOGARCH-EGARCH is stable across $RCov$ frequencies, that is, it only represents the ability of the GOGARCH-EGARCH to mimic the dynamics in the covariance structure generated by the true DGP.

The fact that frequencies associated to the true model seem low when the loss is computed with respect to the true covariance is explained by the fact that we allow for a fairly high degree of similarity between models. The true CCC-EGARCH model with a moderate leverage effect can also be accounted for by other models in the set. However, the presence of the leverage effect in the DGP should imply that all symmetric models are detected as inferior. ¿From Table 3 we also learn that when the quality of the proxy deteriorates (the sampling frequency decreases), the sample performance is invariant, showing the consistency of the ranking of the loss functions across $RCov$ frequencies. The informative content of the loss function is therefore independent from the proxy quality allowing to correctly order the models only on the basis of their forecasting ability.

An interesting case is the CCC-RM. The frequencies associated to the CCC-RM increase progressively from about 8% at $ICov$ to about 10% at $RCov_{12h}$ revealing a behavior that, as we will see in the following, is typically due to the presence of the objective bias. However, the set of models includes also the CCC-IGARCH, a model which shares most of the characteristics of the CCC-RM. The frequencies of the CCC-IGARCH decrease from 5% to 1.5% in such a way to compensate, at each $RCov$ frequency, the increase in the frequency associated to the CCC-RM. The joint probability of CCC-IGARCH and CCC-RM to rank first is indeed about 13% and is constant across $RCov$.

Table 4 reports the same frequencies but for the inconsistent loss functions, i.e. the Frobenius norm, the p-norm with $p = 1$ and the eigenvalue norm. Recall that, when using the true volatility ($ICov$), these loss functions deliver the true ranking. Indeed, the CCC-EGARCH is correctly detected as the best model in 53% of the cases. When relying on $RCov_{1m}$ to $RCov_{1h}$, the frequencies associated to each model remain stable and there is no dominant model other than CCC-EGARCH. Hence, there is no evidence of the presence of objective bias. Starting from $RCov_{2h}$, the frequency at which the CCC-EGARCH model ranks first starts declining while the performance of potentially inferior models increases rapidly as the quality of the proxy lowers. The CCC-EGARCH frequency drops from about 52% to about 38% and 28% at the 12h and daily frequency respectively. Interestingly, for lower levels of proxy quality other inferior models seem to emerge, namely the GOGARCH-EGARCH and the CCC-RM. These models rank first in 18% and about 5% of the cases respectively when using $RCov$ computed from very high frequency data. When using 12h returns to proxy the unobservable covariance (i.e. $RCov_{12h}$) these frequencies increase to about 29% and 20% respectively, meaning that these models rank first quite often. This improvement in the sample performance of these models, as the frequency of $RCov$ lowers, signals the presence of objective bias.

In the first part of the analysis we focused on the detection of the best model in terms of sample performance. However, the analysis carried out so far, offers only a partial insight on the role of the objective bias. Indeed, in presence of a high degree of dissimilarity between the true and the competing models, the detection of the best model may not be affected by the presence of the objective bias. However, the objective bias may still be relevant for what concerns the other

Table 3: Frequencies at which each model shows the smallest sample loss: consistent loss functions

	D-BEKK	RM	CCC-G	CCC-E	CCC-I	CCC-RM	GOG-G	GOG-E	GOG-I	GOG-HY
					Square of the Frobenius norm					
$ICov$	0.002	0.008	0.058	**0.508**	0.052	0.068	0.036	0.172	0.076	0.020
$RCov_{1m}$	0.002	0.006	0.058	**0.510**	0.048	0.068	0.040	0.172	0.072	0.024
$RCov_{5m}$	0.002	0.006	0.060	**0.504**	0.048	0.076	0.042	0.166	0.070	0.026
$RCov_{10m}$	0.004	0.004	0.054	**0.512**	0.040	0.084	0.036	0.168	0.070	0.028
$RCov_{15m}$	0.002	0.008	0.056	**0.504**	0.040	0.076	0.038	0.174	0.076	0.026
$RCov_{20m}$	0.002	0.006	0.048	**0.520**	0.044	0.082	0.036	0.172	0.074	0.016
$RCov_{30m}$	0.002	0.004	0.058	**0.512**	0.042	0.084	0.038	0.170	0.072	0.018
$RCov_{1h}$	0.002	0.002	0.058	**0.522**	0.032	0.092	0.034	0.156	0.072	0.030
$RCov_{2h}$	0.006	0.004	0.048	**0.528**	0.030	0.070	0.034	0.172	0.070	0.038
$RCov_{3h}$	0.002	0.006	0.038	**0.526**	0.036	0.090	0.034	0.152	0.080	0.036
$RCov_{4h}$	0.006	0.002	0.044	**0.496**	0.040	0.098	0.026	0.170	0.074	0.044
$RCov_{6h}$	0.002	0.006	0.042	**0.524**	0.026	0.082	0.022	0.162	0.096	0.038
$RCov_{8h}$	0.006	0.006	0.040	**0.494**	0.018	0.130	0.040	0.152	0.080	0.034
$RCov_{12h}$	0.010	0.000	0.050	**0.526**	0.018	0.100	0.022	0.158	0.078	0.038
$RCov_{1d}$	0.006	0.002	0.036	0.526	0.008	0.130	0.024	0.162	0.066	0.040
					Square of the p-norm with p=1					
$ICov$	0.006	0.012	0.058	**0.470**	0.050	0.094	0.036	0.174	0.082	0.018
$RCov_{1m}$	0.006	0.012	0.054	**0.472**	0.050	0.090	0.036	0.182	0.078	0.020
$RCov_{5m}$	0.006	0.008	0.058	**0.470**	0.050	0.086	0.036	0.178	0.084	0.024
$RCov_{10m}$	0.006	0.006	0.058	**0.464**	0.046	0.098	0.032	0.176	0.086	0.028
$RCov_{15m}$	0.008	0.010	0.062	**0.470**	0.040	0.084	0.038	0.182	0.076	0.030
$RCov_{20m}$	0.004	0.008	0.056	**0.480**	0.050	0.098	0.038	0.162	0.082	0.022
$RCov_{30m}$	0.006	0.004	0.056	**0.484**	0.044	0.102	0.038	0.168	0.074	0.024
$RCov_{1h}$	0.002	0.004	0.060	**0.498**	0.036	0.094	0.034	0.166	0.076	0.030
$RCov_{2h}$	0.010	0.006	0.046	**0.496**	0.038	0.086	0.030	0.174	0.070	0.044
$RCov_{3h}$	0.006	0.008	0.050	**0.494**	0.032	0.104	0.030	0.154	0.082	0.040
$RCov_{4h}$	0.008	0.002	0.048	**0.480**	0.034	0.106	0.030	0.168	0.072	0.052
$RCov_{6h}$	0.008	0.008	0.050	**0.498**	0.022	0.086	0.020	0.164	0.090	0.054
$RCov_{8h}$	0.004	0.006	0.038	**0.462**	0.030	0.146	0.034	0.158	0.082	0.040
$RCov_{12h}$	0.014	0.002	0.048	**0.504**	0.014	0.116	0.020	0.156	0.078	0.048
$RCov_{1d}$	0.004	0.002	0.032	0.498	0.008	0.146	0.026	0.174	0.068	0.042
					Square of the Eigenvalue norm					
$ICov$	0.004	0.008	0.058	**0.502**	0.050	0.072	0.040	0.174	0.072	0.020
$RCov_{1m}$	0.002	0.006	0.056	**0.498**	0.050	0.076	0.036	0.174	0.076	0.026
$RCov_{5m}$	0.004	0.006	0.060	**0.508**	0.046	0.074	0.038	0.166	0.072	0.026
$RCov_{10m}$	0.004	0.008	0.054	**0.504**	0.042	0.084	0.034	0.172	0.072	0.026
$RCov_{15m}$	0.006	0.006	0.058	**0.496**	0.036	0.078	0.038	0.178	0.076	0.028
$RCov_{20m}$	0.004	0.006	0.052	**0.516**	0.044	0.078	0.036	0.168	0.076	0.020
$RCov_{30m}$	0.008	0.006	0.058	**0.504**	0.036	0.086	0.038	0.172	0.074	0.018
$RCov_{1h}$	0.006	0.006	0.054	**0.514**	0.032	0.088	0.036	0.162	0.072	0.030
$RCov_{2h}$	0.006	0.006	0.048	**0.526**	0.026	0.072	0.036	0.178	0.064	0.038
$RCov_{3h}$	0.004	0.006	0.038	**0.524**	0.034	0.090	0.036	0.152	0.080	0.036
$RCov_{4h}$	0.008	0.002	0.046	**0.492**	0.036	0.100	0.030	0.168	0.072	0.046
$RCov_{6h}$	0.006	0.006	0.042	**0.512**	0.026	0.084	0.022	0.166	0.092	0.044
$RCov_{8h}$	0.006	0.006	0.040	**0.500**	0.020	0.128	0.036	0.148	0.078	0.038
$RCov_{12h}$	0.012	0.000	0.052	**0.528**	0.016	0.096	0.020	0.158	0.078	0.040
$RCov_{1d}$	0.006	0.002	0.036	0.522	0.008	0.134	0.026	0.164	0.060	0.042

Note: D-BEKK: Diagonal BEKK; RM: RiskMetrics; CCC-G,-E,-I,-RM: Constant Conditional Correlation with GARCH, EGARCH, IGARCH and Riskmetrics univariate conditional variances; GOG-G,-E,-I,-HY: Generalized Orthogonal GARCH with GARCH, EGARCH, IGARCH and HYGARCH univariate conditional variances. $RCov_{1d}$ is separated by a horizontal line indicating that the realized covariance matrix is not positive definite at the daily frequency.

Table 4: Frequencies at which each model shows the smallest sample loss: inconsistent loss functions

	D-BEKK	RM	CCC-G	CCC-E	CCC-I	CCC-RM	GOG-G	GOG-E	GOG-I	GOG-HY
					Frobenius norm					
$ICov$	0.000	0.002	0.046	**0.528**	0.034	0.050	0.030	0.182	0.106	0.022
$RCov_{1m}$	0.000	0.002	0.048	**0.526**	0.032	0.050	0.030	0.182	0.108	0.022
$RCov_{5m}$	0.000	0.002	0.048	**0.526**	0.026	0.056	0.034	0.180	0.108	0.020
$RCov_{10m}$	0.000	0.002	0.046	**0.528**	0.026	0.058	0.032	0.182	0.104	0.022
$RCov_{15m}$	0.000	0.002	0.044	**0.536**	0.028	0.052	0.030	0.180	0.104	0.024
$RCov_{20m}$	0.000	0.002	0.044	**0.534**	0.024	0.054	0.028	0.184	0.108	0.022
$RCov_{30m}$	0.000	0.002	0.044	**0.528**	0.022	0.058	0.028	0.180	0.110	0.028
$RCov_{1h}$	0.000	0.002	0.038	**0.528**	0.028	0.072	0.022	0.188	0.096	0.026
$RCov_{2h}$	0.000	0.002	0.046	**0.502**	0.032	0.084	0.012	0.214	0.084	0.024
$RCov_{3h}$	0.000	0.004	0.032	**0.508**	0.024	0.094	0.014	0.206	0.094	0.024
$RCov_{4h}$	0.002	0.006	0.038	**0.480**	0.028	0.110	0.010	0.216	0.084	0.026
$RCov_{6h}$	0.006	0.016	0.018	**0.444**	0.022	0.132	0.010	0.256	0.078	0.018
$RCov_{8h}$	0.002	0.020	0.020	**0.422**	0.018	0.170	0.004	0.256	0.072	0.016
$RCov_{12h}$	0.010	0.026	0.012	**0.392**	0.010	0.202	0.000	0.290	0.050	0.008
$RCov_{1d}$	0.012	0.048	0.006	0.294	0.004	0.288	0.002	0.298	0.042	0.006
					p-norm with p=1					
$ICov$	0.004	0.004	0.054	**0.506**	0.024	0.060	0.030	0.186	0.110	0.022
$RCov_{1m}$	0.004	0.004	0.052	**0.504**	0.028	0.060	0.032	0.182	0.110	0.024
$RCov_{5m}$	0.004	0.004	0.046	**0.506**	0.024	0.066	0.038	0.180	0.108	0.024
$RCov_{10m}$	0.006	0.004	0.042	**0.516**	0.024	0.064	0.034	0.174	0.108	0.028
$RCov_{15m}$	0.006	0.002	0.040	**0.512**	0.022	0.068	0.030	0.180	0.114	0.026
$RCov_{20m}$	0.006	0.002	0.040	**0.506**	0.024	0.072	0.032	0.178	0.116	0.024
$RCov_{30m}$	0.006	0.002	0.044	**0.506**	0.022	0.068	0.028	0.184	0.114	0.026
$RCov_{1h}$	0.008	0.002	0.046	**0.504**	0.024	0.076	0.018	0.190	0.106	0.026
$RCov_{2h}$	0.010	0.002	0.052	**0.476**	0.022	0.096	0.010	0.218	0.090	0.024
$RCov_{3h}$	0.010	0.008	0.042	**0.474**	0.022	0.100	0.016	0.212	0.092	0.024
$RCov_{4h}$	0.012	0.008	0.030	**0.458**	0.022	0.122	0.010	0.224	0.088	0.026
$RCov_{6h}$	0.016	0.018	0.012	**0.424**	0.020	0.150	0.008	0.258	0.070	0.024
$RCov_{8h}$	0.018	0.028	0.012	**0.402**	0.010	0.178	0.008	0.260	0.068	0.016
$RCov_{12h}$	0.024	0.028	0.006	**0.376**	0.010	0.208	0.000	0.292	0.052	0.004
$RCov_{1d}$	0.022	0.052	0.004	0.272	0.002	0.298	0.002	0.298	0.046	0.004
					Eigenvalue norm					
$ICov$	0.002	0.002	0.050	**0.520**	0.032	0.050	0.032	0.180	0.110	0.022
$RCov_{1m}$	0.000	0.002	0.052	**0.516**	0.032	0.054	0.030	0.184	0.106	0.024
$RCov_{5m}$	0.002	0.002	0.046	**0.518**	0.028	0.058	0.036	0.182	0.108	0.020
$RCov_{10m}$	0.000	0.002	0.042	**0.522**	0.026	0.058	0.032	0.182	0.112	0.024
$RCov_{15m}$	0.000	0.002	0.044	**0.534**	0.026	0.056	0.030	0.178	0.104	0.026
$RCov_{20m}$	0.002	0.002	0.052	**0.520**	0.022	0.054	0.028	0.178	0.120	0.022
$RCov_{30m}$	0.002	0.002	0.044	**0.514**	0.022	0.060	0.026	0.184	0.114	0.032
$RCov_{1h}$	0.002	0.002	0.042	**0.526**	0.026	0.070	0.022	0.188	0.092	0.030
$RCov_{2h}$	0.000	0.002	0.046	**0.506**	0.028	0.082	0.014	0.216	0.080	0.026
$RCov_{3h}$	0.004	0.004	0.036	**0.498**	0.022	0.094	0.016	0.206	0.096	0.024
$RCov_{4h}$	0.004	0.008	0.040	**0.472**	0.024	0.108	0.014	0.222	0.084	0.024
$RCov_{6h}$	0.008	0.020	0.016	**0.444**	0.020	0.130	0.012	0.248	0.082	0.020
$RCov_{8h}$	0.010	0.024	0.020	**0.422**	0.012	0.162	0.002	0.260	0.070	0.018
$RCov_{12h}$	0.014	0.028	0.010	**0.394**	0.010	0.198	0.000	0.288	0.050	0.008
$RCov_{1d}$	0.016	0.052	0.004	0.296	0.004	0.280	0.002	0.298	0.042	0.006

Note: D-BEKK: Diagonal BEKK; RM: RiskMetrics; CCC-G,-E,-I,-RM: Constant Conditional Correlation with GARCH, EGARCH, IGARCH and Riskmetrics univariate conditional variances; GOG-G,-E,-I,-HY: Generalized Orthogonal GARCH with GARCH, EGARCH, IGARCH and HYGARCH univariate conditional variances. $RCov_{1d}$ is separated by a horizontal line indicating that the realized covariance matrix is not positive definite at the daily frequency.

positions in the ranking. We now investigate whether the whole ordering is preserved despite the deterioration of the quality of the proxy. Since we are ranking over a set of estimated volatility models, the true ranking, except for the best model, is not known ex ante. However the underlying ordering implied by a given loss function, can be identified by ranking the models with respect to the true covariance $ICov$.

A general result appears from Tables 3 and 4. Unsurprisingly, the frequencies reported are homogeneous between loss functions within each group. This is a direct result of the equivalences stated in (5) and (20), for the consistent, (11) and (21) for the inconsistent loss functions. Hence, without loss of generality, we consider next only one consistent (Frobenius norm squared) and one inconsistent (Frobenius norm) loss function. Figure 1(a) shows the ranking based on the average performance (over the 500 replications) implied by the consistent loss function for various levels of proxy quality. This ranking is fairly stable across $RCov$ frequencies meaning that the squared Frobenius norm is able to consistently order models even when the quality of the proxy deteriorates. Shifts in position affect only the middle of the classification and can be justified by the extremely close average sample performances between the models, with differences at $RCov_{1d}$ smaller than 10^{-2} (Figure 1(b)). Figures 1(b) and 1(c) provide some insights to disentangle the role of the accuracy of the covariance proxy. Figure 1(c) reports models average performances normalized to the average performance of the CCC-EGARCH model. The converging pattern towards the true model, together with constant deviations between models across $RCov$ frequencies (Figure 1(b)) suggests that the loss of accuracy only translates into a increase in the level of the average sample performances for all models. Constant discrepancies between models imply that not only the ordering but also the degree of similarity, and therefore the relationships between models, are preserved across frequencies. However, the increase in the variability of the proxy induces an increase in the variability of the loss function which, in empirical applications, may result in the impossibility to effectively discriminate between models.

A different picture emerges when considering the inconsistent loss function (Figure 2(a)). In this case, the ranking is preserved up to one hour sampling frequency. Due to the appearance of the objective bias, we observe major shifts at lower frequencies at most levels of the classification. The impact of the objective bias is amplified by the fact that except for the first two positions, i.e. CCC-EGARCH and GOGARCH-EGARCH, all the other models exhibit very close average sample performances (Figure 2(b)), with differences smaller than 10^{-2} at $RCov_{1d}$. Inferior models like RiskMetrics and CCC-RM, 10th and 9th respectively according to $ICov$, improve up to the 3rd and 2nd positions respectively. The CCC-EGARCH is classified as the best forecasting model at all frequencies, followed by the GOGARCH-EGARCH. This result is due to the fact that these two models are sufficiently different from the others (they are the only models in the set to allow for leverage effect of the same type as implied by the true model), with the CCC-EGARCH clearly dominating the GOGARCH-EGARCH (Figure 2(b)). Although the objective bias does not become an issue when ordering between these two models, Figure 2(b) shows that, as the frequency for $RCov$ lowers, the average sample performance of the latter gets closer to the CCC-EGARCH performance. Since, as underlined above, the variability of the loss function increases along with the variability of the proxy, the probability to rank the GOGARCH-EGARCH first increases at low frequences. This conclusion is consistent with the results reported in Table 4.

Besides varying the proxy quality and studying several loss functions we also investigate the impact of the estimation sample size on the rankings. We increase the sample size by 50% to 3000 observations and we find qualitatively identical results (results available on request).

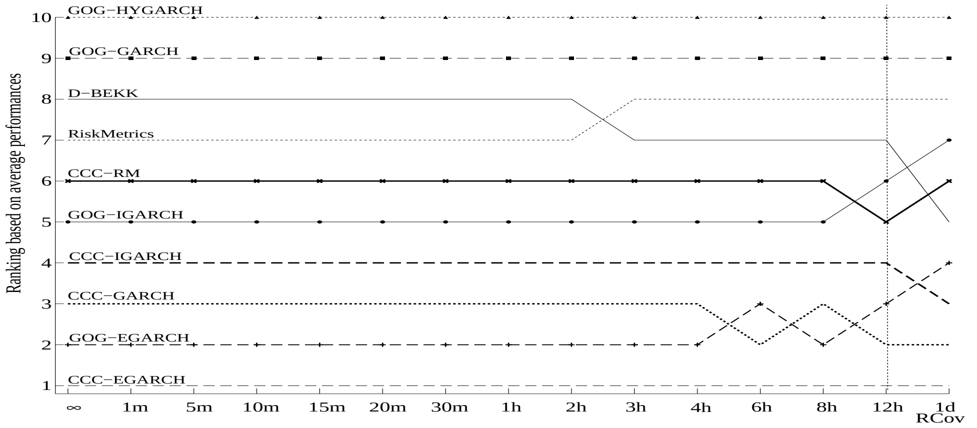
6 Empirical application

6.1 Data description and estimation results

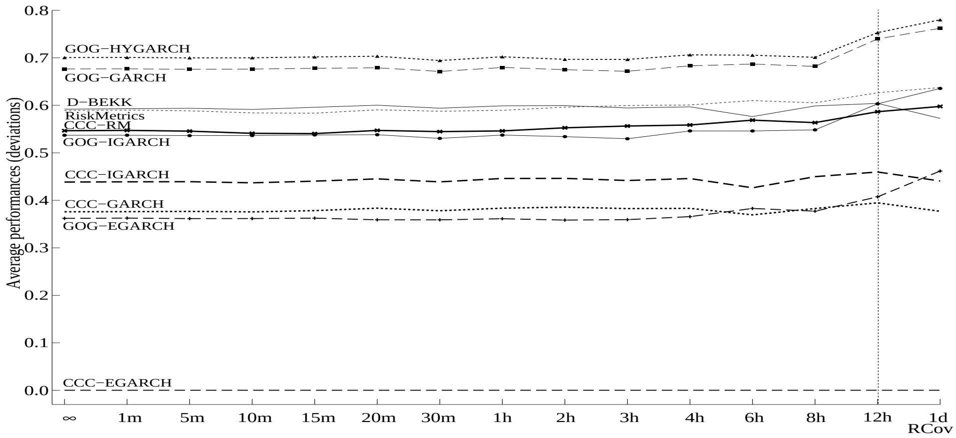
The empirical application is based on the Euro, British Pound and the Japanese Yen exchange rates expressed in US dollars (EUR, GBP and JPY). The sample period is January 6, 1987 through June 30, 2004 totaling 4287 trading days. Intraday returns and realized covariances are computed from five-minutes intervals last mid-quotes, implying 288 intraday observations. The data have been provided by Olsen & Associates. Missing values are replaced by linearly interpolating the closest previous and the first following 5-minute price. The dataset has been cleaned from weekends, holidays and early closing days. Days with too many missing values and/or constant prices are also removed. Five-minute returns are computed as the first difference of the logarithmic prices. The estimation sample ranges from January 6, 1987 to December 28, 2001 (3666 trading days), while the remaining observations (621 trading days) are used for the out-of-sample forecasts evaluation. Table 5 reports descriptive statistics for the estimation sample and the forecasting sample. With respect to the daily frequency, the EUR and GBP exchange rates share similar data characteristics and are relatively highly correlated. JPY has quite a higher kurtosis and a more pronounced skewness. The 5-minute realized variances and correlations are quite dispersed. For example the correlations vary between -0.12 and 0.85. We also remark that the variances are positively skewed and the correlations negatively skewed.

The proxy for the conditional covariance is realized covariance ($RCov$) as defined in (27) computed at 14 different frequencies ranging from 5 min. to 24 h. We stress again, like in the simulation study, that we should stop at the 8h frequency if we want to have a positive definite realized variance matrix at each point in time. We include the results until the 24h frequency to illustrate what happens when the realized variance matrix is not positive definite. One-step-ahead forecasts are computed from 4:05 pm to 4:00 pm ET and are contrasted to the realized measure of

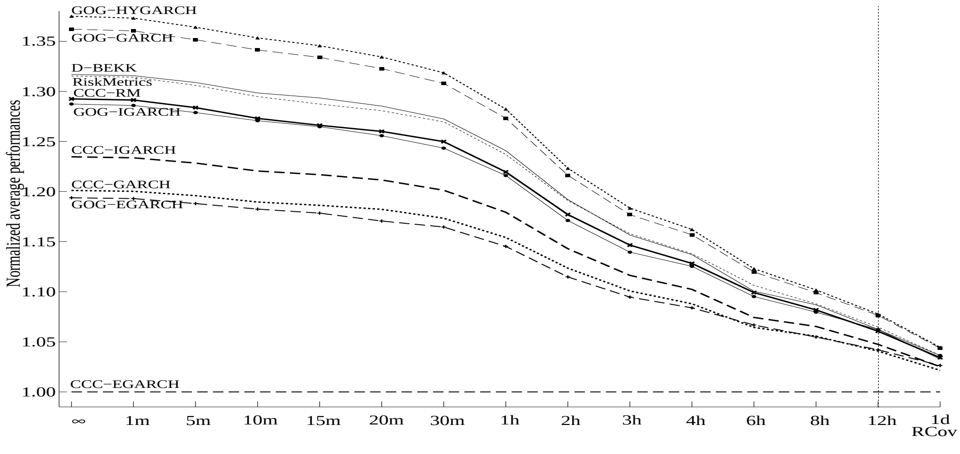
Figure 1: Consistency of the ranking based on average performance - Frobenius norm squared



(a) Ranking based on average performances



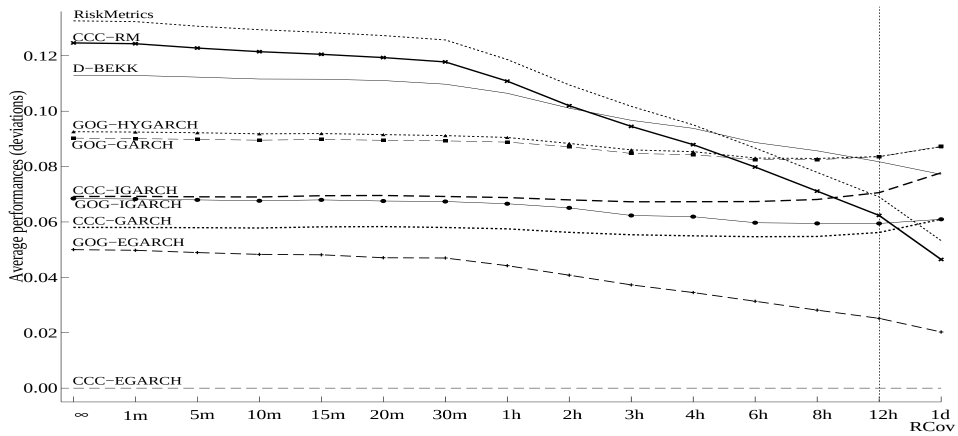
(b) Avg. performances - Deviations from the CCC-EGARCH



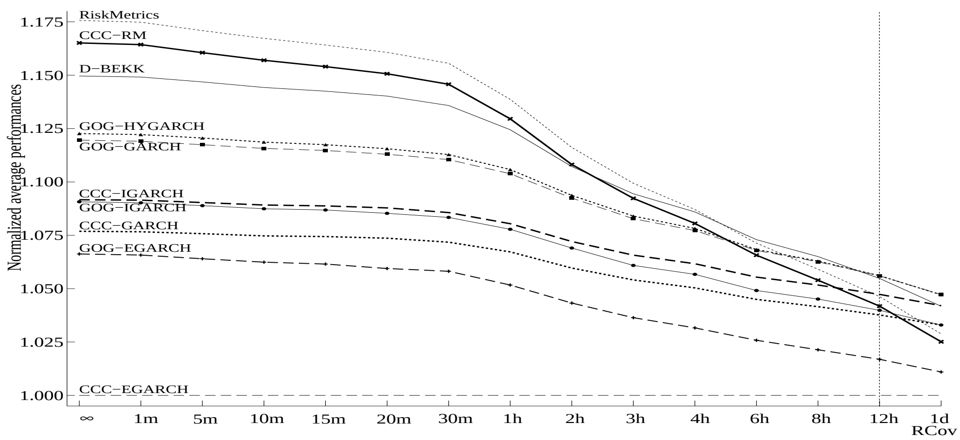
(c) Normalized average performances

Figure 2: Consistency of the ranking based on average performance - Frobenius norm

(a) Ranking based on average performances



(b) Avg. performances - Deviations from the CCC-EGARCH



(c) Normalized average performances

Table 5: Descriptive statistics

Series	min	mean	max	std.dev	skew.	kurt.
Estimation sample: January 6, 1987 to December 28, 2001 (3666 observations)						
EUR	-3.557	-0.003	3.419	0.683	0.043	4.939
GBP	-4.168	-0.002	3.425	0.623	-0.161	6.140
JPY	-4.207	0.003	7.724	0.729	0.619	9.503
Forecasting sample: January 3, 2002 to June 30, 2004 (621 observations)						
EUR	-2.001	0.051	1.837	0.647	-0.227	3.270
GBP	-1.756	0.035	2.051	0.524	-0.221	3.873
JPY	-2.203	0.033	2.686	0.595	-0.129	4.260
$RCov_{5m,EUR}$	0.122	0.457	2.526	0.200	3.024	24.52
$RCov_{5m,GBP}$	0.079	0.315	1.564	0.156	2.410	14.02
$RCov_{5m,JPY}$	0.041	0.413	2.385	0.235	3.221	20.52
$RCor_{5m,EUR,GBP}$	0.012	0.550	0.852	0.120	-0.303	3.359
$RCor_{5m,EUR,JPY}$	-0.035	0.410	0.800	0.147	-0.343	2.639
$RCor_{5m,GBP,JPY}$	-0.122	0.279	0.653	0.127	-0.131	2.885

The estimated correlations for the estimation sample are $\rho_{EUR,GBP} = 0.720$, $\rho_{EUR,JPY} = 0.493$ and $\rho_{GBP,JPY} = 0.415$. The estimated correlations for the forecasting sample are $\rho_{EUR,GBP} = 0.721$, $\rho_{EUR,JPY} = 0.490$ and $\rho_{GBP,JPY} = 0.416$.

volatility using one consistent (Frobenius norm squared) and one inconsistent (Frobenius norm) loss function. Estimation results for the 16 MGARCH models are reported in Table 6. Note that there is no Riskmetrics and CCC-RM procedures reported in Table 6 since they do not require parameter estimation (the sample correlation is used for the CCC-RM). Generally speaking, we observe that the parameters estimates for the conditional variance, covariance and correlations imply highly persistent processes. Furthermore, in almost all cases, the null of no leverage effect cannot be rejected at standard significance levels.

6.2 Model comparison

The empirical ranking of the 16 MGARCH models, as a function of the level of aggregation of the data used to compute $RCov$, is reported in Figures 3 and 4. The consistent loss function in Figure 3(a) points to the CCC-GARCH as the best forecasting model at almost all frequencies. More generally we can conclude that the subset given by both the CCC and the DCC both with GARCH and GJR outperforms all the other models. This model is followed by the CCC-GJR, with differences between the two rather negligible (Figure 3(b)). The overall ranking is well preserved across all frequencies.

The GOGARCH model is always largely dominated by all other models regardless the conditional variance specification. There is no clear dominance between the CCC and the DCC models and their ranking position depends on the model chosen for the conditional variance. Here the GARCH/GJR represents the best combination, followed in the order by the APARCH, the RM and finally the IGARCH. Interestingly, the three models which are based on the RiskMetrics approach, which assumes dynamics for the variance process independent of the data by fixing a smoothing parameter ex ante (RiskMetrics, CCC-RM and DCC-RM) are positioned in the middle of the classification. Figure 3(a) shows that between 10 min. and 1 h., the ranking is particularly stable but rather volatile outside this range of frequencies. The accuracy of the volatility proxy plays an important role here. As pointed out by Hansen and Lunde (2006) we can observe discrepancies between the empirical and the approximated ranking in finite samples (i.e. sampling error). Indeed, as the accuracy of the proxy deteriorates, the loss function becomes less informative. As a result, it is more difficult to identify superior models. This effect becomes more severe when there is a high degree of similarity between models under evaluation. The relationship between intraday frequency and accuracy has been discussed in Section 4 and 5. However, we underline that $RCov$ may also be severely biased when based on intraday returns sampled at very high frequency, due to the presence of jumps and microstructure noise.

Figure 4(a) illustrates how the presence of the objective bias can affect the ranking when an inconsistent loss function is used. The overall ordering between models is generally preserved and stable across frequencies with three striking exception. The CCC and the DCC models with RM

Table 6: Estimation results

	DBEKK		*CCC*				*DCC*					*GOG*			
			G	*I*	*A*	*J*	*G*	*I*	*A*	*J*		*G*	*I*	*A*	*J*
$\hat{c}_{11}$	0.060	$\hat{\rho}_{21}$	0.710	0.729	0.710	0.710	0.706	0.758	0.707	0.708	$\hat{\varphi}_1$	0.086	0.084	0.111	0.090
	(0.008)		(0.011)	(0.010)	(0.011)	(0.011)	(0.055)	(0.057)	(0.053)	(0.054)		(0.058)	(0.055)	(0.005)	(0.059)
$\hat{c}_{21}$	0.044	$\hat{\rho}_{31}$	0.511	0.545	0.518	0.511	0.643	0.732	0.613	0.645	$\hat{\varphi}_2$	0.179	0.185	0.167	0.182
	(0.007)		(0.016)	(0.015)	(0.015)	(0.016)	(0.064)	(0.059)	(0.067)	(0.064)		(0.044)	(0.050)	(0.012)	(0.044)
$\hat{c}_{13}$	0.039	$\hat{\rho}_{32}$	0.430	0.462	0.432	0.430	0.511	0.608	0.480	0.511	$\hat{\varphi}_3$	0.417	0.431	0.376	0.416
	(0.007)		(0.017)	(0.017)	(0.017)	(0.017)	(0.076)	(0.073)	(0.077)	(0.076)		(0.086)	(0.088)	(0.008)	(0.087)
$\hat{c}_{22}$	0.041	$\hat{\vartheta}_p$					0.026	0.026	0.025	0.025	$\hat{c}$	0.006	0.002	0.006	0.007
	(0.006)						(0.003)	(0.003)	(0.003)	(0.003)		(0.003)	(0.001)	(0.004)	(0.004)
$\hat{c}_{23}$	0.006	$\hat{\vartheta}_q$					0.971	0.971	0.971	0.971	$\hat{\alpha}$	0.033	0.035	0.033	0.041
	(0.004)						(0.004)	(0.004)	(0.004)	(0.004)		(0.009)	(0.008)	(0.011)	(0.013)
$\hat{c}_{33}$	0.050	$\hat{c}$	0.013	0.003	0.018	0.014	0.007	0.002	0.008	0.007	$\hat{\beta}$	0.961		0.959	0.959
	(0.008)		(0.004)	(0.001)	(0.005)	(0.004)	(0.002)	(0.001)	(0.002)	(0.002)		(0.010)		(0.011)	(0.012)
$\hat{a}_{11}$	0.192	$\hat{\alpha}$	0.034	0.044	0.045	0.033	0.040	0.044	0.048	0.045	$\hat{\gamma}$			−0.102	−0.013
	(0.011)		(0.006)	(0.006)	(0.010)	(0.006)	(0.005)	(0.006)	(0.007)	(0.006)				(0.087)	(0.011)
$\hat{a}_{22}$	0.201	$\hat{\beta}$	0.936		0.933	0.934	0.945		0.943	0.945	$\hat{\delta}$			2.078	
	(0.017)		(0.011)		(0.013)	(0.012)	(0.008)		(0.059)	(0.009)				(0.419)	
$\hat{a}_{22}$	0.188	$\hat{\gamma}$			0.032	0.005			−0.074	−0.009	$\hat{c}$	0.007	0.005	0.012	0.008
	(0.014)				(0.083)	(0.009)			(0.059)	(0.007)		(0.004)	(0.003)	(0.005)	(0.005)
		$\hat{\delta}$			1.314				1.586		$\hat{\alpha}$	0.065	0.068	0.077	0.060
					(0.171)				(0.214)			(0.020)	(0.022)	(0.018)	(0.018)
$\hat{g}_{11}$	0.978	$\hat{c}$	0.010	0.003	0.012	0.012	0.004	0.002	0.003	0.004	$\hat{\beta}$	0.930		0.928	0.927
	(0.003)		(0.004)	(0.001)	(0.007)	(0.005)	(0.002)	(0.001)	(0.002)	(0.002)		(0.022)		(0.076)	(0.023)
$\hat{g}_{22}$	0.976	$\hat{\alpha}$	0.037	0.045	0.037	0.022	0.040	0.042	0.031	0.032	$\hat{\gamma}$			0.160	0.014
	(0.004)		(0.010)	(0.010)	(0.014)	(0.009)	(0.009)	(0.009)	(0.013)	(0.009)				(0.076)	(0.016)
$\hat{g}_{22}$	0.979	$\hat{\beta}$	0.933		0.930	0.930	0.952		0.954	0.953	$\hat{\delta}$			0.909	
	(0.003)		(0.019)		(0.023)	(0.021)	(0.012)		(0.012)	(0.012)				(0.217)	
		$\hat{\gamma}$			0.194	0.028			0.068	0.011	$\hat{c}$	0.009	0.005	0.007	0.008
					(0.082)	(0.013)			(0.062)	(0.010)		(0.004)	(0.002)	(0.004)	(0.004)
		$\hat{\delta}$			1.900				2.306		$\hat{\alpha}$	0.050	0.055	0.040	0.038
					(0.311)				(0.391)			(0.012)	(0.013)	(0.013)	(0.011)
		$\hat{c}$	0.009	0.003	0.011	0.009	0.009	0.003	0.012	0.010	$\hat{\beta}$	0.941		0.945	0.944
			(0.004)	(0.001)	(0.006)	(0.004)	(0.004)	(0.001)	(0.006)	(0.005)		(0.014)		(0.013)	(0.014)
		$\hat{\alpha}$	0.045	0.049	0.054	0.050	0.045	0.047	0.058	0.052	$\hat{\gamma}$			0.096	0.019
			(0.012)	(0.012)	(0.013)	(0.016)	(0.010)	(0.009)	(0.013)	(0.015)				(0.058)	(0.013)
		$\hat{\beta}$	0.939		0.946	0.938	0.937		0.937	0.936	$\hat{\delta}$			2.309	
			(0.018)		(0.015)	(0.019)	(0.004)		(0.018)	(0.019)				(0.378)	
		$\hat{\gamma}$			−0.225	−0.009			−0.128	−0.013					
					(0.116)	(0.013)			(0.086)	(0.011)					
		$\hat{\delta}$			0.797				1.269						
					(0.186)				(0.268)						
L	−8481		−9007	−9085	−8975	−8999	−8469	−8498	−8452	−8463		−8643	−8653	−8623	−8637

Notes: DCC-RM: $\hat{\rho}_{21} = 0.576$ (0.061), $\hat{\rho}_{31} = 0.461$ (0.075), $\hat{\rho}_{32} = 0.300$ (0.094), $\hat{\vartheta}_p = 0.028$ (0.002) and $\hat{\vartheta}_q = 0.969$ (0.002).
L=loglikelihood value at the maximum likelihood estimates. Standard errors between parentheses. Definitions of the models are in Table 1.

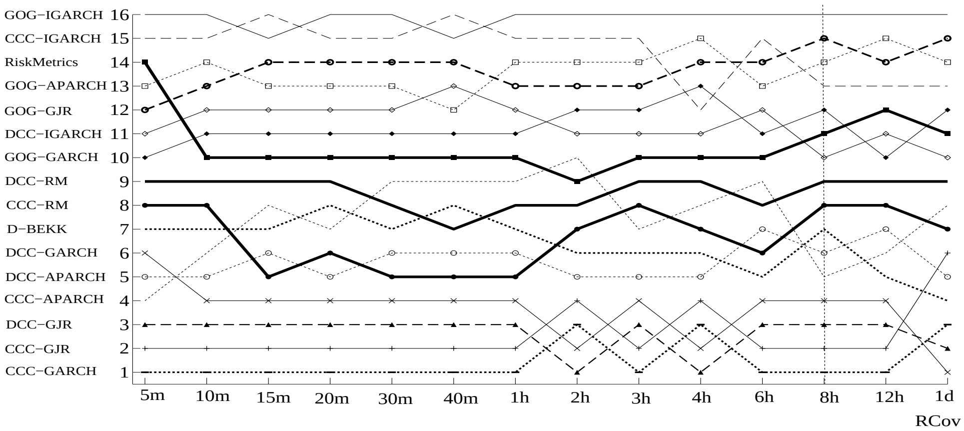
conditional variances rank 8th and 9th respectively at $RCov_{5min}$, but they rapidly climb towards the top of the classification as the frequency for $RCov$ lowers. Starting from 15min frequency for $RCov$ they reach the top of the classification, ranking first and second. Interestingly, (Figure 4(b)), the sample performances of these two models are extremely close, with discrepancies from the average across models at each frequency ranging between 0 and 0.003. Similarly, the RiskMetrics model, ranking 11th when $RCov_{5m}$ is used, joins the top of the ranking at a relatively high frequency. When $RCov$ is computed using data sampled at a frequency equal or lower then 40 minutes, the RiskMetrics model ranks 3rd, behind the CCC-RM and DCC-RM model. Given that these models are characterized by a dynamic in the variance structure imposed ex ante and independent from the data (with the only exception of the DCC-RM for which the parameters of the dynamic correlation are data dependent), it is unlikely that such models are the best forecasting model. The presence of a biased ordering is therefore striking. The ranking obtained at low frequencies is in no way compatible with the one obtained when a more accurate proxy is used. Since model performances are extremely close (Figure 4(b)), the objective bias severely affects the ranking even when the proxy used in the evaluation is based on rather high frequency data.

In Figures 5 and 6, we concentrate the analysis on a reduced set of models that includes only non nested models (CCC is excluded), since the CCC and the DCC models are rather equivalent in terms of sample performances. Since we consider models characterized by a lower degree of similarity, the impact of sampling error is now reduced. The ranking implied by the consistent loss function is highly stable for a larger range of frequencies. Again, when the inconsistent loss function is used the appearance of the objective bias clearly affects the ordering. In Figure 6(b), we observe the relative improvement in terms of sample performances of the DCC-RM and the RiskMetrics models with respect to all the others models in the set, with a striking dominance of the DCC-RM.

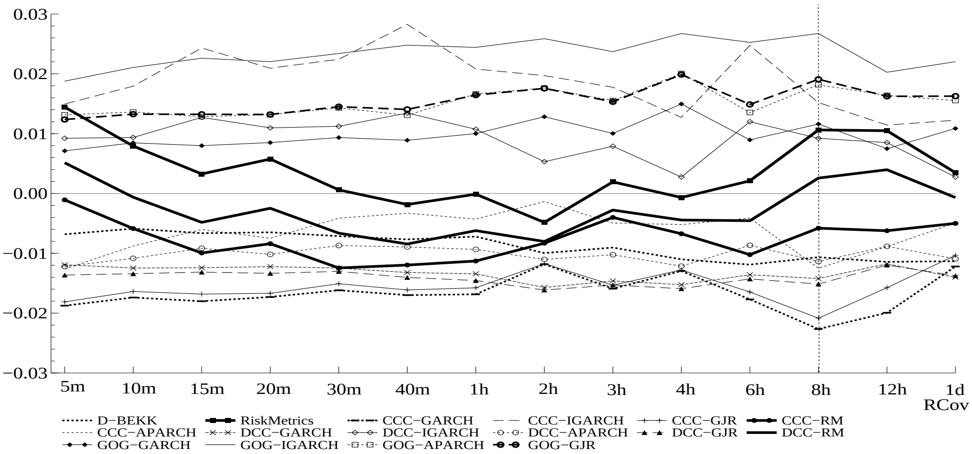
6.3 Model confidence set

To illustrate the crucial role of an adequate choice of the loss function for model selection based on forecasting ability, we apply the Model Confidence Set (MCS) test of Hansen, Lunde, and Nason (2005) to the set of models with lower degree of similarity. The MCS test allows to identify a subset of equivalent models in terms of predictive ability which are superior to the others. Being the selection based on the ordering implied by the loss function used to evaluate the deviations from the target volatility (i.e. the implied orderings shown in Figures 5(a) and 6(a)), an unfortunate choice of the loss function can deliver an unintended result even when the testing procedure is formally correct. Table 7 summarizes the results for three different sampling frequencies for the covariance proxy $RCov$.

Figure 3: Consistency of the ranking based on sample performances - Consistent loss function

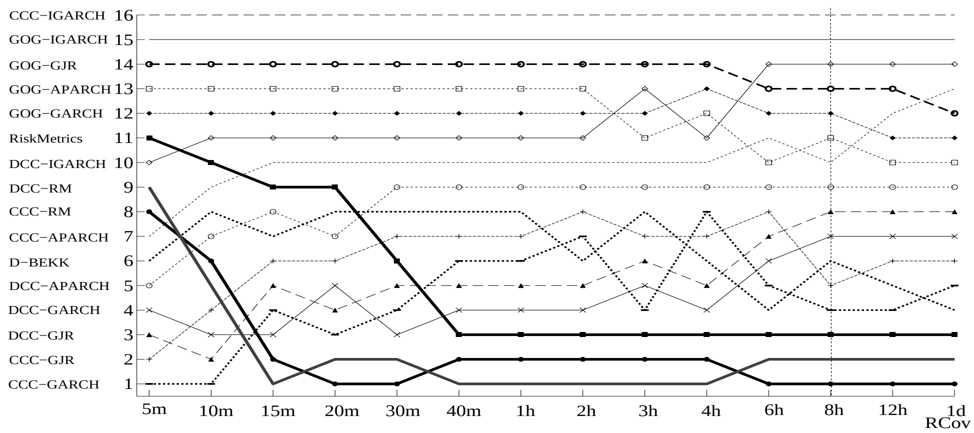


(a) Frobenius norm squared - Ranking

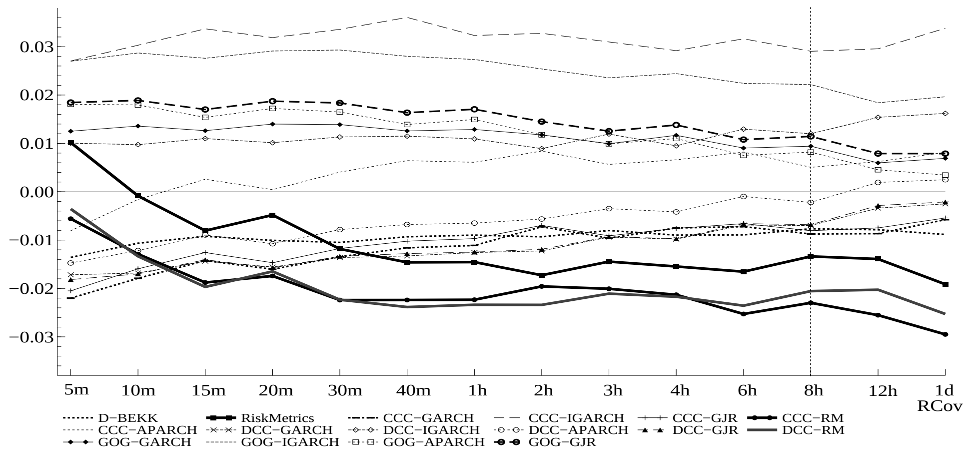


(b) Frobenius norm squared - Deviations from the average across models

Figure 4: Consistency of the ranking based on sample performances - Inconsistent loss function

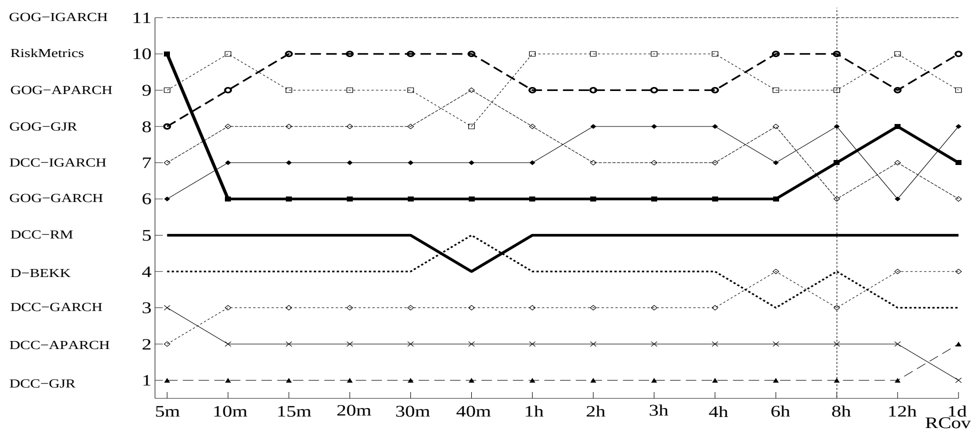


(a) Frobenius norm - Ranking

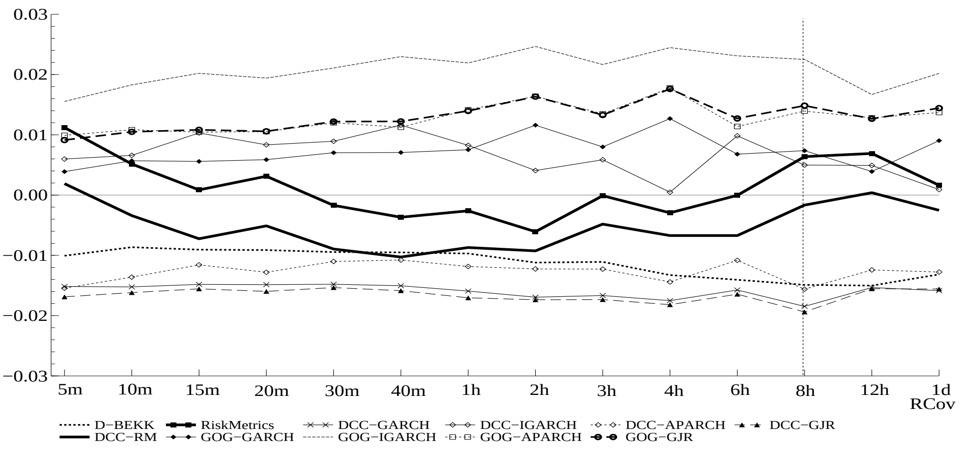


(b) Frobenius norm - Deviations from the average across models

Figure 5: Consistency of the ranking based on sample performances (reduced set) - Consistent loss function

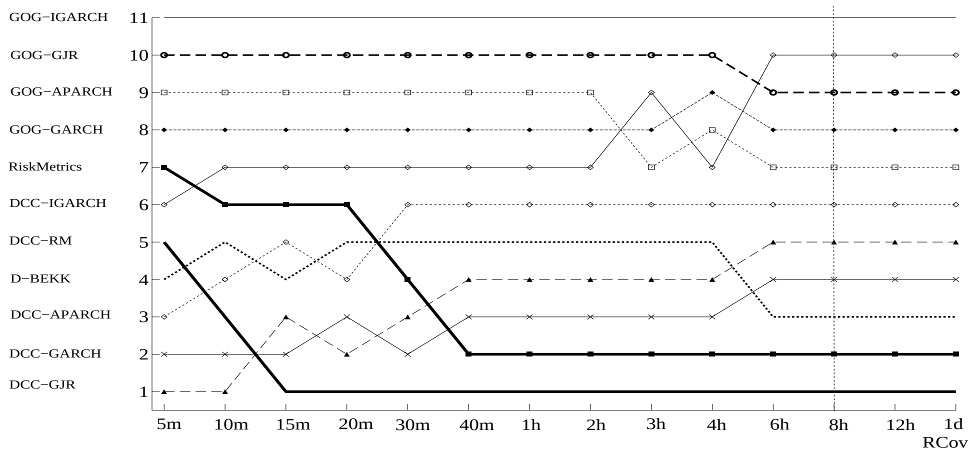


(a) Frobenius norm squared - Ranking

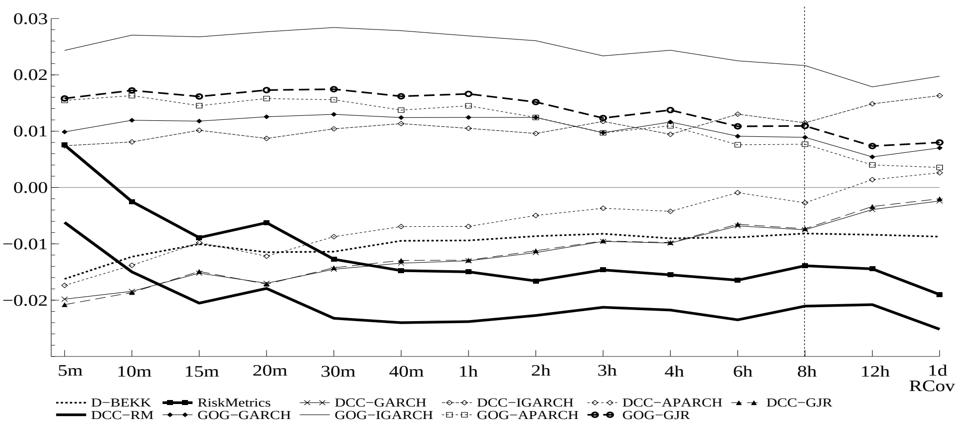


(b) Frobenius norm squared - Deviations from the average across models

Figure 6: Consistency of the ranking based on sample performances (reduced set) - Inconsistent loss function



(a) Frobenius norm - Ranking



(b) Frobenius norm - Deviations from the average across models

We apply first the MCS test using the Frobenius norm squared (consistent). The results reported in Table 7 are consistent across frequencies. Furthermore, the set of equally good models gets larger as the sampling frequency for $RCov$ lowers. This result is due to the loss of accuracy of the proxy which translates in a higher variability of the sample evaluation associated to each model, which consequently makes more difficult to discriminate between models.

Table 7: Model Confidence Set test.

Loss function	$RCov_{5m}$	$RCov_{20m}$	$RCov_{8h}$
Frobenius norm squared	DCC-APARCH DCC-GARCH DCC-GJR	DCC-GARCH DCC-GJR	DCC-APARCH DCC-GARCH DCC-GJR D-BEKK
Frobenius norm	DCC-GARCH DCC-GJR	DCC-GARCH DCC-GJR DCC-RM	DCC-RM

Notes: The initial set contains 11 models. Significance level $\alpha = 0.05$.
Sample size 621 obs. Standard errors based on 1000 bootstrap resamples.

When the test is based on the inconsistent Frobenius norm, the result appears clearly affected by the objective bias. In fact, the MCS gets smaller and its composition changes as the frequency for $RCov$ lowers. At $RCov_{8h}$ the set is made of the only DCC-RM. Indeed, this model shows a largely better performance with respect to all other models, as shown in Figure 6(b).

7 Conclusion

Two important issues arise when we want to rank several multivariate volatility models with respect to their forecasting performance. First, there is the choice of the loss function (how can we compare predicted variance matrices) and second the choice of a proxy of the unobservable volatility measure used to evaluate models forecasts. In fact, when the unobservable volatility is substituted by a proxy, the ordering implied by some loss functions may be biased with respect to the intended one.

In this paper, we extend Hansen and Lunde (2006) conditions for consistent ranking to the multivariate case. Interestingly, it turns out that, being in a multivariate framework, we can put forward a new condition for consistent ranking based on norm equivalence that broadens the class of admissible loss functions. We discuss in this paper several loss functions which are based on transformations of existing matrix norms and verify whether they satisfy the conditions to ensure a consistent ranking. The proxy of the unobservable volatility matrix is the realized covariance matrix.

In the simulation study, we sample from a continuous time multivariate diffusion process and estimate discrete time multivariate GARCH models to illustrate the sensitivity of the ranking to different choices of the loss functions and to the quality of the proxy. We observe that if the quality of the proxy is good, both consistent and inconsistent loss functions rank properly. However, when the quality of the proxy is poor, only the consistent loss functions rank properly. Our findings also hold when the sample size in the estimation period increases. This is an important message for the applied econometrician.

The application to three foreign exchange rates nicely illustrates, with respect to losses and a model confidence set test, what happens when we pick a poor proxy combined with an inconsistent loss function in an out-of-sample forecast comparison among 16 multivariate GARCH models. We actually observe that for the foreign exchange rates series the models perform similarly in predicting conditional variance matrices.

There are interesting extensions for future research. First, this paper ranks multivariate volatility models based on statistical loss functions only and focuses on conditions for consistent ranking from a more theoretical viewpoint. At some point an economic loss function has to be introduced when the forecasted volatility matrices are actually used in financial applications such as portfolio management and option pricing. It is clear that the model with the smallest statistical loss is always preferred but it may happen that other models with small statistical loss have economic loss properties that are indistinguishable. This issue has not been addressed in this paper. Second, multivariate volatility forecast comparison for higher horizons than one day is not studied yet. Third, other proxies than realized covariance that enter the loss functions should be investigated.

References

Andersen, T., T. Bollerslev, P. Christoffersen, and F. Diebold (2006): *Volatility and Correlation Forecasting* chap. 15, pp. 777–877, Handbook of Economic Forecasting. North Holland.

Andersen, T., T. Bollerslev, F. Diebold, and P. Labys (2003): “Modeling and Forecasting Realized Volatility,” *Econometrica*, 71, 579–625.

Barndorff-Nielsen, O., and N. Shephard (2004): “Measuring the impact of jumps in multivariate price processes using bipower covariation,” *DP, Nuffield College, Oxford University*, 2, 1–37.

Bauwens, L., S. Laurent, and J. Rombouts (2006): “Multivariate GARCH Models: A Survey,” *Journal of Applied Econometrics*, 21, 79–109.

Bollerslev, T. (1986): "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 307–327.

——— (1990): "Modeling the Coherence in Short-run Nominal Exchange Rates: A Multivariate Generalized ARCH model," *Review of Economics and Statistics*, 72, 498–505.

Clark, T., and M. McCracken (2001): "Tests of Equal Forecast Accuracy and Encompassing for Nested Models," *Journal of Econometrics*, 105, 85–110.

Davidson, J. (2004): "Moment and Memory Properties of Linear Conditional Heteroscedasticity Models, and a New Model," *Journal of Business and Economic Statistics*, 22, 16–29.

Diebold, F., and R. Mariano (1995): "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics*, 13, 253–263.

Ding, Z., C. W. J. Granger, and R. F. Engle (1993): "A Long Memory Property of Stock Market Returns and a New Model," *Journal of Empirical Finance*, 1, 83–106.

Elliott, G., and A. Timmermann (2008): "Economic Forecasting," *Journal of Economic Literature*, 46, 3–56.

Engle, R. (2002): "Dynamic Conditional Correlation - a Simple Class of Multivariate GARCH Models," *Journal of Business and Economic Statistics*, 20, 339–350.

Engle, R., and T. Bollerslev (1986): "Modelling the Persistence of Conditional Variances," *Econometric Reviews*, 5, 1–50.

Engle, R., and F. Kroner (1995): "Multivariate Simultaneous Generalized ARCH," *Econometric Theory*, 11, 122–150.

Glosten, L., R. Jagannathan, and D. Runkle (1992): "On the Relation Between the Expected Value and Volatility of the Nominal Excess Return on Stocks," *Journal of Finance*, 46, 1779–1801.

Golub, G., and C. Van Loan (1996): *Matrix Computations*. The Johns Hopkins University Press.

Hafner, C. (2007): "Temporal Aggregation of Multivariate GARCH Processes," *Journal of Econometrics*, 142, 467–483.

Hansen, P., and A. Lunde (2005): "A Forecast Comparison of Volatility Models: Does Anything Beat a GARCH(1,1)," *Journal of Applied Econometrics*, 20, 873–889.

——— (2006): "Consistent Ranking of Volatility Models," *Journal of Econometrics*, 131, 97–121.

Hansen, P., A. Lunde, and J. Nason (2003): "Choosing the Best Volatility Models: The Model Confidence Set Approach," Federal Reserve of Atlanta Working Paper.

——— (2005): "Model Confidence Sets for Forecasting Models," Federal Reserve of Atlanta Working Paper.

Horn, R., and C. Johnson (1985): *Matrix Analysis.* Cambridge University Press.

J.P.Morgan (1996): *Riskmetrics Technical Document, 4th ed.* J.P.Morgan, New York.

Laurent, S. (2007): *G@RCH 5. Estimating and Forecasting Garch Models.* Timberlake Consultants Ltd.

McAller, M., and M. Medeiros (2008): "Realized Volatility: A Review," *Econometric Reviews*, 27.

Meddahi, N. (2002): "A Theoretical Comparison between Integrated and Realized Volatility," Université de Montréal Working paper.

Nelson, D. (1991): "Conditional Heteroskedasticity in Asset Returns: a New Approach," *Econometrica*, 59, 349–370.

van der Weide, R. (2002): "GO-GARCH: A Multivariate Generalized Orthogonal GARCH Model," *Journal of Applied Econometrics*, 17, 549–564.

Voev, V., and A. Lunde (2006): "Integrated Covariance Estimation Using High Frequency Data in Presence of Noise," *Journal of Financial Econometrics*, 5.

West, K. (1996): "Asymptotic Inference About Predictive Ability," *Econometrica*, 64, 1067–1084.

White, H. (2000): "Reality Check for Data Snooping," *Econometrica*, 68, 1097–1126.